Baker- Lin-Huang type Bivariate distributions based on order statistics

I. Bairamov[1] and K. Bayramoglu[2]

[1]Department of Mathematics Izmir University of Economics, Izmir, Turkey.
E-mail: ismihan.bayramoglu@ieu.edu.tr

[2]Department of Statistics, Middle East Technical University, Ankara,Turkey.
E-mail: konul@metu.edu.tr

Abstract

Baker (2008) introduced a new class of bivariate distributions based on distributions of order statistics from two independent samples of size n. Lin-Huang (2010) discovered an important property of Baker's distribution and showed that the Pearson's correlation coefficient for this distribution converges to maximum attainable value, i.e. the correlation coefficient of the Frechét upper bound, as n increases to infinity. Bairamov and Bayramoglu (2011) investigated a new class of bivariate distributions constructed by using Baker's model and distributions of order statistics from dependent random variables, allowing high correlation than that of Baker's distribution. In this paper a new class of Baker's type bivariate distributions with high correlation are constructed on the base of distributions of order statistics by using an arbitrary continuous copula instead of the product copula.

Keywords: Bivariate distribution function, FGM distributions, copula, positive quadrant dependent, negative quadrant dependent, order statistics, Pearson's correlation coefficient.

1 Introduction

Huang and Kotz (1999) introduced new modifications of classical Farlie-Gumbel-Morgenstern (FGM) distribution introducing additional parameters. The new Huang-Kotz FGM distributions allow correlation higher than the classical FGM and because of the simple analytical form aroused interest of many researchers. Last years there appeared many papers dealing with the modifications of FGM distribution allowing high correlation. For a related works on this subject see Bairamov et al. (2001), Amblard and Girard (2002), Bairamov and Kotz (2002),

(2003), Fisher and Klein (2007) among others. Baker (1998) use a novel approach connected with the FGM distribution and introduce a new class of bivariate distributions based on the distributions of order statistics. Baker (2008) considers independent random variables X_i, Y_i, from two univariate distributions with distribution functions (c.d.f.) F_X and F_Y, respectively. The corresponding probability density functions (p.d.f.) are f_X and f_Y. Let $U_1 = \min(X_1, X_2)$ and $U_2 = \max(X_1, X_2)$, $V_1 = \min(Y_1, Y_2)$ and $V_2 = \max(Y_1, Y_2)$. To obtain positive correlation Baker randomly chooses either the pair U_1, V_1 or U_2, V_2. The random numbers are now positively correlated, but the marginal distributions are unchanged, because the random choice of either of two order statistics from a distribution gives a random variable from that distribution. To obtain a negative correlation, either U_1, V_2 or U_2, V_1 chosen. The bivariate distribution of a randomly chosen pair of order statistics is

$$1/2 \times \{F_X^{2:2}(x)F_Y^{2:2}(y) + F_X^{1:2}(x)F_Y^{1.2}(y)\}$$

and on choosing either a pair of order statistics with probability q or the original independent random variables X, Y with probability $1 - q$, the bivariate distribution function is

$$
\begin{aligned}
H(x,y) &= (1-q)F_X(x)F_Y(y) + (q/2)\{F_X^{2:2}(x)F_Y^{2:2}(y) + F_X^{1:2}(x)F_Y^{1.2}(y)\} \\
&= F_X(x)F_Y(y)\{1 + q(1 - F_X(x))(1 - F_Y(y))\}.
\end{aligned}
$$

In general let $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ...Y_n$ are independent and identically distributed (i.i.d.) random variables with distribution functions (d.f.) F_X and F_Y, respectively. Let $X_{k:n}$ and $Y_{k:n}$, $k = 1, 2, ..., n$ be corresponding order statistics and $F_X^{k:n}(x) = P\{X_{k:n} \leq x\}$, $F_Y^{k:n}(y) = P\{Y_{k:n} \leq y\}$. Baker's bivariate distribution function is now defined as

$$
H_+^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x)F_Y^{k:n}(y) \tag{1}
$$

$$
H_-^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x)F_Y^{n-k+1:n}(y). \tag{2}
$$

For Baker's bivariate distribution $H_+^{(n)}(x,y)$ with exponential marginals $F_X(x) = F_Y(x) = 1 - e^{-x}, x > 0$, the Pearson's correlation coefficient is $\rho_n = 1 - \frac{1}{n}\sum_{k=1}^{n}\frac{1}{k}$, which increases monotonely to 1.

As a generalization of (1) and (2) Baker also introduces

$$
H_r^{(n)}(x,y) = \sum_{k=1}^{n}\sum_{l=1}^{n} r_{kl} F_X^{k:n}(x)F_Y^{l:n}(y), \tag{3}
$$

where $r_{kl} \geq 0$ and $\sum_{k=1}^{n} r_{kl} = \sum_{l=1}^{n} r_{kl} = \frac{1}{n}$, for all $k, l = 1, 2, ..., n$. Lin and Huang (2010) proved that (3) does not contain members with a correlation higher than that of (1). That is why the best bivariate distribution

with higher positive correlation among the members (3) is (1). Similar consideration holds true for the negative correlation. Lin and Huang (2010) investigate the conditions under which the correlation for (1) converges to the limit. In particular, they show that if either (i) $X \geq b, Y \geq c$ a.s. for some $b, c \in \mathbb{R}$ and $E(X_{k:n}) \geq F_X^{-1}(\frac{k-1}{n})$ and $E(Y_{k:n}) \geq F_Y^{-1}(\frac{k-1}{n})$ for all (k, n), or (ii) $X \leq b, Y \leq c$ a.s. for some $b, c \in \mathbb{R}$ and $E(X_{k:n}) \leq F_X^{-1}(\frac{k}{n})$ and $E(Y_{k:n}) \leq F_Y^{-1}(\frac{k}{n})$ for all (k, n) then $\lim_{n \to \infty} \rho_n = \rho^*$, where ρ^* is the correlation coefficient of the Fréchet-Hoeffding upper bound $H_+(x, y) = \min(F_X(x), F_Y(y))$ (see Frechéet (1940)). The results presented in the paper of Lin and Huang (2010) makes Baker's distribution attractive.

Recently, Bairamov and Bayramoglu (2011) observed that if in the Baker's model instead of independent random variables one uses the dependent random variables (X, Y) with positive quadrant dependent (PQD) joint distribution function $F(x, y)$, then the correlation increases and for negative quadrant dependent $F(x, y)$, decreases. More precisely, let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be a bivariate sample with joint distribution function $F(x, y) = C(F_X(x), F_Y(y))$. Bairamov and Bayramoglu (2011) consider the following bivariate distribution functions constructed on the basis of the Baker's idea:

$$K_+^{(n)}(x, y) = \frac{1}{n} \sum_{r=1}^{n} P\{X_{r:n} \leq x, Y_{r:n} \leq y\}$$

$$K_-^{(n)}(x, y) = \frac{1}{n} \sum_{r=1}^{n} P\{X_{r:n} \leq x, Y_{n-r+1:n} \leq y\},$$

where $X_{i:n}$ and $Y_{j:n}$ are the ith and jth order statistics constructed on the basis of bivariate observations (X_i, Y_i), $(i = 1, 2, ..., n)$ with joint distribution function $F(x, y) = P\{X_i \leq x, Y_i \leq y\}$ and marginal distribution functions $F_X(x) = F(x, \infty)$, $F_Y(y) = F(\infty, y)$ so that $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$, $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$. The joint d.f. of $X_{r:n}$ and $Y_{s:n}$ is given in David (1981) (see also Arnold et al. (1992)) as

$$\begin{aligned} P\{X_{r:n} &\leq x, Y_{s:n} \leq y\} \\ &= \sum_{i=r}^{n} \sum_{j=s}^{n} \sum_{k=a}^{b} c(n, k; i, j) p_{11}^{k} p_{12}^{i-k} p_{21}^{j-k} p_{22}^{n-i-j+k}, \end{aligned} \quad (4)$$

where

$$c(n, k; i, j) = \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!}, \ a = \max(0, i+j-n), \ b = \min(i, j) \quad (5)$$

and

$$\begin{aligned} p_{11} &= F(x, y) \\ p_{12} &= F_X(x) - F(x, y) \\ p_{21} &= F_Y(y) - F(x, y) \\ p_{22} &= 1 - F_X(x) - F_Y(y) + F(x, y). \end{aligned} \quad (6)$$

3

Then

$$K_+^{(n)}(x,y) \quad = \quad \frac{1}{n}\sum_{r=1}^{n}\sum_{i=r}^{n}\sum_{j=r}^{n}\sum_{k=a}^{b} c(n,k;i,j)p_{11}^{k}p_{12}^{i-k}p_{21}^{j-k}p_{22}^{n-i-j+k} \qquad (7)$$

$$K_-^{(n)}(x,y) \quad = \quad \frac{1}{n}\sum_{r=1}^{n}\sum_{i=r}^{n}\sum_{j=n-r+1}^{n}\sum_{k=a}^{b} c(n,k;i,j)p_{11}^{k}p_{12}^{i-k}p_{21}^{j-k}p_{22}^{n-i-j+k}. \qquad (8)$$

It is clear that the marginal distributions of $K_+^{(n)}(x,y)$ and $K_-^{(n)}(x,y)$ are again $F_X(x)$ and $F_Y(y)$, i.e. $K_+^{(n)}(x,\infty) = F_X(x)$ and $K_+^{(n)}(\infty,y) = F_Y(y)$. It is shown that for a PQD joint distribution function $F(x,y)$ the positive correlation of $K_+^{(n)}(x,y)$ is higher than that of $H_+^{(n)}(x,y)$ and for NQD $F(x,y)$ and the negative correlation of $K_-^{(n)}(x,y)$ is smaller than that of $H_-^{(n)}(x,y)$.

In this note we consider a new class of bivariate distribution functions using Baker's construction and considering any copula $C(u,v)$ instead of product copula $C(u,v) = \Pi(u,v) = uv$. It follows that for this new class of distributions if $C(u,v)$ is PQD, i.e. $C(u,v) \geq uv$, for all $(u,v) \in [0,1]^2$, then the Pearson's correlation coefficient is higher than that of Baker's distribution. Similarly, if the copula is NQD, i.e. $C(u,v) \leq uv$, for all $(u,v) \in [0,1]^2$ then then the Pearson's correlation coefficient is smaller than that of Baker's distribution. All modifications constructed on the Baker's idea we call as Baker-Lin-Huang Type distributions. The distribution (7), (8) we call Baker's Type I BB and the new distributions introduced in this paper Baker's Type II BB distributions.

2 New bivariate Baker's Type II BB distributions based on a copula approach

Let $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ...Y_n$ be i.i.d. random variables with d.f.'s F_X and F_Y, respectively. Let $X_{k:n}$ and $Y_{k:n}$, $k = 1, 2, ..., n$ be corresponding order statistics and $F_X^{k:n}(x) = P\{X_{k:n} \leq x\}$, $F_Y^{k:n}(y) = P\{Y_{k:n} \leq y\}$.

Recall that a two-dimensional copula is a function $C(x,y)$ from $[0,1]^2 = [0,1] \times [0,1]$ to $[0,1]$ with the properties:

1. $C(x,0) = 0 = C(0,y)$, $C(x,1) = x$ and $C(1,y) = y$;
2. For every x_1, x_2, y_1, y_2 such that $0 \leq x_1 < x_2 \leq 1$ and $0 \leq y_1 < y_2 \leq 1$

$$C(x_2,y_2) - C(x_2,y_1) - C(x_1,y_2) + C(x_1,y_1) \geq 0.$$

According to Sklar's Theorem if $F_{X,Y}(x,y)$ is a joint distribution function with continuous marginal distributions $F_X(x)$ and $F_Y(y)$, then there exists a unique copula C such that $F(x,y) = C(F_X(x), F_Y(y))$. Theory and applications of cop-

ulas are well documented in Nelsen (2005). Let $C(u,v)$ be any copula. Consider

$$G_+^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} C(F_X^{k:n}(x), F_Y^{k:n}(y)) \tag{9}$$

$$G_-^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} C(F_X^{k:n}(x), F_Y^{n-k+1:n}(y)). \tag{10}$$

From the properties of a copula it follows that the marginal distributions of $G_+^{(n)}(x,y)$ and $G_-^{(n)}(x,y)$ are F_X and F_Y, respectively. In fact,

$$G_+^{(n)}(x,\infty) = \frac{1}{n}\sum_{k=1}^{n} C(F_X^{k:n}(x),1) = \frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x) = F_X(x)$$

$$G_+^{(n)}(\infty,y) = \frac{1}{n}\sum_{k=1}^{n} C(1,F_Y^{k:n}(y)) = \frac{1}{n}\sum_{k=1}^{n} F_Y^{k:n}(y) = F_Y(y).$$

Similarly, $G_-^{(n)}(x,\infty) = F_X(x)$ and $G_-^{(n)}(\infty,y) = F_Y(y)$. Hereafter, we will denote by ρ_H the Pearson's correlation coefficient between any random variables X and Y with joint distribution function $H(x,y)$. It is clear that if $C(u,v) = \Pi(u,v) = uv$ then $G_+^{(n)}(x,y) = H_+^{(n)}(x,y)$ and $G_-^{(n)}(x,y) = H_-^{(n)}(x,y)$. Since $C(F_X^{k:n}(x), F_Y^{k:n}(y))$ is a bivariate c.d.f. (with marginals $F_X^{k:n}(x)$ and $F_Y^{k:n}(y)$) then $G_+^{(n)}(x,y)$ is obviously a bivariate c.d.f. as a convex combination of bivariate c.d.f.'s. The copula used in construction (9) and (10) will be called the "kernel" copula for Baker's Type II BB distribution.

Theorem 1. If $C(u,v)$ is PQD then $\rho_{G_+^{(n)}} \geq \rho_{H_+^{(n)}}$ and if $C(u,v)$ is NQD then $\rho_{G_-^{(n)}} \leq \rho_{H_-^{(n)}}$.

Proof. Since $C(u,v) \geq uv$ for all $(u,v) \in [0,1]^2$, then $C(F_X^{k:n}(x), F_Y^{k:n}(y)) \geq F_X^{k:n}(x)F_Y^{k:n}(y)$ for all $(x,y) \in \mathbb{R}^2$. From the Hoeffding's formula (see Hoeffding

(1940)) for correlation coefficient one has

$$\rho_{G_+^{(n)}} = \frac{1}{\sqrt{Var(x)Var(Y)}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} [G_+^{(n)}(x,y) - F_X(x)F_Y(y)]dxdy$$

$$= \frac{1}{\sqrt{Var(x)Var(Y)}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} [\frac{1}{n}\sum_{k=1}^{n} C(F_X^{k:n}(x), F_Y^{k:n}(y))$$

$$-F_X(x)F_Y(y)]dxdy \tag{11}$$

$$\rho_{H_+^{(n)}} = \frac{1}{\sqrt{Var(x)Var(Y)}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} [H_+^{(n)}(x,y) - F_X(x)F_Y(y)]dxdy$$

$$= \frac{1}{\sqrt{Var(x)Var(Y)}} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} [\frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x)F_Y^{k:n}(y)$$

$$-F_X(x)F_Y(y)]dxdy \tag{12}$$

and $\rho_{G_+^{(n)}} \geq \rho_{H_+^{(n)}}$. Similarly, $\rho_{G_-^{(n)}} \leq \rho_{H_-^{(n)}}$.

Example 1. Baker's Type II BB distributions with FGM "kernel" copula. Let

$$C(u,v) = uv(1 + \alpha(1-u)(1-v)), (u,v) \in [0,1]^2, -1 \leq \alpha \leq 1. \tag{13}$$

Consider

$$G_+^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x)F_Y^{k:n}(y)(1+ \tag{14}$$

$$\alpha(1 - F_X^{k:n}(x))(1 - F_Y^{k:n}(y)), \ 0 \ \leq \ \alpha \leq 1.$$

$$G_-^{(n)}(x,y) = \frac{1}{n}\sum_{k=1}^{n} F_X^{k:n}(x)F_Y^{n-k+1:n}(y)(1+$$

$$\alpha(1 - F_X^{k:n}(x))(1 - F_Y^{n-k+1:n}(y)), \ -1 \ \leq \ \alpha \leq 0. \tag{15}$$

Let $F_X(x) = x, 0 \leq x \leq 1$ and $F_Y(y) = y, 0 \leq y \leq 1$. Since the FGM copula (13) is PQD for $\alpha \geq 0$ and is NQD for $\alpha \leq 0$, then $\rho_{G_+^{(n)}} \geq \rho_{H_+^{(n)}}$ for $\alpha \geq 0$, $\rho_{G_-^{(n)}} \leq \rho_{H_-^{(n)}}$, for $\alpha \leq 0$. It is clear that if $\alpha = 0$ then $G_+^{(n)}(x,y) = H_+^{(n)}(x,y)$ and $G_-^{(n)}(x,y) = H_-^{(n)}(x,y)$. In the following table we present some numerical values of $\rho_{G_+^{(n)}}, \rho_{K_+^{(n)}}, \rho_{H_+^{(n)}}, \rho_{G_-^{(n)}}, \rho_{K_-^{(n)}}$ and $\rho_{H_-^{(n)}}$ for $Uniform(0,1)$ marginals. The numerical calculations are made in MATLAB which is one of the commonly accepted packages for coding mathematical models since it has built-in functions for probability distributions and allows probabilistic and mathematical operations.

6

n	2	4	6	8	10	12	15	20
$\rho_{G_+^{(n)}}$	0.5467	0.7258	0.8039	0.8475	0.8753	0.8945	0.9144	0.9348
$\rho_{K_+^{(n)}}$	0.5133	0.6915	0.7761	0.8247	0.8561	0.8779	0.9006	0.9241
$\rho_{H_+^{(n)}}$	0.3333	0.6000	0.7143	0.7778	0.8182	0.8462	0.8750	0.9048
$\rho_{G_-^{(n)}}$	-0.5467	-0.7258	-0.8039	-0.8475	-0.8753	-0.8945	-0.9144	-0.9348
$\rho_{K_-^{(n)}}$	-0.5133	-0.6915	-0.7761	-0.8247	-0.8561	-0.8779	-0.9006	-0.9241
$\rho_{H_-^{(n)}}$	-0.3333	-0.6000	-0.7143	-0.7778	-0.8182	-0.8462	-0.8750	-0.9048

Table 1. Correlation coefficients $\rho_{G_+^{(n)}}$ and $\rho_{G_-^{(n)}}$ with

$$F(x,y) \;=\; xy(1 + \alpha(1-x)(1-y)), \alpha = 1, F_X(x) = x, F_Y(y) = y, 0 \le x, y \le 1$$
$$\text{and } \rho_{K_+^{(n)}}, \rho_{K_-^{(n)}} \text{ and } \rho_{H_+^{(n)}}, \rho_{H_-^{(n)}} \text{ with } Uniform(0,1) \text{ marginals.}$$

It can be observed from the Table 1 that $\rho_{G_-^{(n)}} \le \rho_{K_-^{(n)}} \le \rho_{H_-^{(n)}} \le \rho_{H_+^{(n)}} \le \rho_{K_+^{(n)}} \le \rho_{G_+^{(n)}}$.

In the following table the values of the correlation coefficients for $\rho_{G_+^{(n)}}, \rho_{K_+^{(n)}} \rho_{H_+^{(n)}}$, $\rho_{G_-^{(n)}}, \rho_{K_-^{(n)}}$ and $\rho_{H_-^{(n)}}$ for $Uniform(0,1)$ and $Exponential(1)$ marginal distributions.

n	2	4	6	8	10	12	15	20
$\rho_{G_+^{(n)}}$	0.4811	0.6343	0.7001	0.7367	0.7600	0.7762	0.7929	0.8102
$\rho_{K_+^{(n)}}$	0.4426	0.5951	0.6682	0.7105	0.7379	0.7572	0.7772	0.7980
$\rho_{H_+^{(n)}}$	0.2886	0.5196	0.6185	0.6735	0.7085	0.7327	0.7577	0.7835
$\rho_{G_-^{(n)}}$	-0.4811	-0.6343	-0.7001	-0.7367	-0.7600	-0.7762	-0.7929	-0.8102
$\rho_{K_-^{(n)}}$	-0.4426	-0.5951	-0.6682	-0.7105	-0.7379	-0.7572	-0.7772	-0.7980
$\rho_{H_-^{(n)}}$	-0.2886	-0.5196	-0.6185	-0.6735	-0.7085	-0.7327	-0.7577	-0.7835

Table 2. Correlation coefficients $\rho_{G_+^{(n)}}$ and $\rho_{G_-^{(n)}}$ with

$$F(x,y) \;=\; xy(1 + \alpha(1-x)(1-y)), \alpha = 1, F_X(x) = x, 0 \le x \le 1, F_Y(y) = 1 - \exp(-y), y \ge 0$$
$$\text{and } \rho_{K_+^{(n)}}, \rho_{K_-^{(n)}} \text{ and } \rho_{H_+^{(n)}}, \rho_{H_-^{(n)}} \text{ with the same marginals.}$$

Again, from the Table 2 we have $\rho_{G_-^{(n)}} \le \rho_{K_-^{(n)}} \le \rho_{H_-^{(n)}} \le \rho_{H_+^{(n)}} \le \rho_{K_+^{(n)}} \le \rho_{G_+^{(n)}}$.

3 Copula representation of Baker's Type II BB distribution

Denote by $\Pi(t, s) = ts$, $(t, s) \in [0, 1]^2$, a product copula. Let (X_i, Y_i), $i = 1, 2, ..n$ be a bivariate sample with joint distribution function $F(x, y) = C(F_X(x), F_Y(y))$. Consider the joint distribution of order statistics $X_{r:n}$ and $Y_{s:n}$ given in (4):

$$
\begin{aligned}
F_{X_{r:n}, Y_{s:n}}(x, y) &= P\{X_{r:n} \leq x, Y_{s:n} \leq y\} \\
&= \sum_{i=r}^{n} \sum_{j=s}^{n} \sum_{k=a}^{b} c(n, k; i, j) p_{11}^{k} p_{12}^{i-k} p_{21}^{j-k} p_{22}^{n-i-j+k}, \quad (16)
\end{aligned}
$$

where $c(n, k; i, j)$, $p_{11}, p_{12}, p_{21}, p_{22}$ are given in (5),(6). The copula of vivariate distribution $F_{X_{r:n}, Y_{s:n}}(x, y)$ presents an interest. Denote this copula as $C_{r,s:n}(t, s)$, then

$$
F_{X_{r:n}, Y_{r:n}}(x, y) = C_{r,s:n}(F_{X_{r:n}}(x), F_{Y_{s:n}}(y)). \quad (17)
$$

It is well known (David (1981))

$$
\begin{aligned}
F_{X_{r:n}}(x) &= \sum_{i=r}^{n} \binom{n}{i} F_X^i(x)(1 - F_X(x))^{n-i} \\
&= \frac{1}{B(r, n-r+1)} \int_{0}^{F_X(x)} u^{r-1}(1-u)^{n-r} du = I_{r,n-r+1}(F_X(x))
\end{aligned}
$$

and

$$
\begin{aligned}
F_{Y_{s:n}}(y) &= \sum_{i=s}^{n} \binom{n}{i} F_Y^i(y)(1 - F_Y(y))^{n-i} \\
&= \frac{1}{B(s, n-s+1)} \int_{0}^{F_Y(y)} u^{s-1}(1-u)^{n-s} du = I_{s,n-s+1}(F_Y(y)),
\end{aligned}
$$

where $I_{a,b}(p) = \frac{1}{B(a,b)} \int_{0}^{p} u^{a-1}(1-u)^{b-1} du$ is an incomplete Beta function. Denote by $I_{a,b}^{-1}(p)$ the inverse of $I_{a,b}(p)$. Let $F_{X_{r:n}}(x) = I_{r,n-r+1}(F_X(x)) = t$ and $F_{Y_{s:n}}(y) = I_{s,n-s+1}(F_Y(y)) = s$. Then

$$
x = F_X^{-1}(I_{r,n-r+1}^{-1}(F_X(t))) \quad \text{and} \quad y = F_Y^{-1}(I_{s,n-s+1}^{-1}(F_Y(s))). \quad (18)
$$

From (17) and (18) one has

$$
C_{r,s:n}(t, s) = F_{X_{r:n}, Y_{r:n}}\left(F_X^{-1}(I_{r,n-r+1}^{-1}(t)), F_Y^{-1}(I_{s,n-s+1}^{-1}(s))\right). \quad (19)
$$

Therefore a copula of joint distribution of order statistics $X_{r:n}$ and $Y_{s:n}$ is $C_{r,s:n}(t,s)$ given in (19). In a special case if $r = s = n$ one has

$$\begin{aligned} C_{n,n:n}(t,s) &= F^n\left(F_X^{-1}(I_{n,1}^{-1}(t)), F_Y^{-1}(I_{n,1}^{-1}(s))\right) \qquad (20) \\ &= F^n(F_X^{-1}(t^{1/n}), F_Y^{-1}(s^{1/n})), \end{aligned}$$

since $I_{n,1}(t) = t^n$ and $I_{n,1}^{-1}(v) = v^{1/n}$.

Analogously, one obtains the copula of joint distribution of order statistics $X_{1:n}$ and $Y_{1:n}$ as

$$\begin{aligned} C_{1,n:n}(t,s) = \ & t + s - 1 \\ &+ [(1-t)^{1/n} + (1-s)^{1/n} - 1 \qquad (21) \\ &+ F(F_X^{-1}(1 - (1-t)^{1/n}), F_Y(1 - (1-s)^{1/n}))]^n \end{aligned}$$

by noting that the joint distribution function of $X_{1:n}$ and $Y_{1:n}$ is

$$\begin{aligned} F_{X_{1:n}, Y_{1:n}}(x,y) &= P\{X_{1:n} \le x, Y_{1:n} \le y\} \\ &= (1 - (1 - F_X(x))^n) + (1 - (1 - F_Y(y))^n) - 1 + \bar{F}^n(x,y) \\ &= (1 - (1 - F_X(x))^n) + (1 - (1 - F_Y(y))^n) - 1 \\ &\quad + (1 - F_X(x) - F_Y(y) + F(x,y))^n \\ &= C_{1,n:n}\left(1 - (1 - F_X(x)))^n, 1 - (1 - F_Y(y))^n\right). \qquad (22) \end{aligned}$$

3.1 The Baker's Type II BB distributions with "kernel" copula of bivariate FGM extreme order statistics

If the underlying distribution is classical FGM, i.e.

$$F(x,y) = F_X(x) F_Y(y)(1 + \alpha(1 - F_X(x))(1 - F_Y(y))), -1 \le \alpha \le 1. \qquad (23)$$

one obtains from (20)

$$C_{n:n}(t,s) = ts\left[1 + \alpha(1 - t^{1/n})(1 - s^{1/n})\right]^n \qquad (24)$$

and from (21)

$$\begin{aligned} C_{1:n}(t,s) = \ & t + s - 1 + \left[(1-t)^{1/n} + (1-s)^{1/n} - 1 \right. \\ &+ (1 - (1-t)^{1/n})(1 - (1-s)^{1/n}) \\ &\left. \times \{1 + \alpha(1-t)^{1/n}(1-s)^{1/n}\}\right]^n. \qquad (25) \end{aligned}$$

It follows that $\lim_{n\to\infty} C_{n:n}(t,s) = ts = \Pi(t,s)$ and $\lim_{n\to\infty} C_{1:n}(t,s) = \Pi(t,s)$,

since $\lim\limits_{n\to\infty}\left[1+\alpha(1-t^{1/n})(1-s^{1/n})\right]^{n}=1$ and

$$
\begin{aligned}
&\lim_{n\to\infty}\Big[(1-t)^{1/n}+(1-s)^{1/n}-1\\
&\quad+(1-(1-t)^{1/n})(1-(1-s)^{1/n})\{1+\alpha(1-t)^{1/n}(1-s)^{1/n}\}\Big]^{n}\\
&=\;(1-t)(1-s),
\end{aligned}
$$

where $\Pi(t,s)$ is a product copula of independent random variables. This means that if the joint distribution of (X,Y) is FGM given in (23), then the extreme order statistics $X_{n:n}$ and $Y_{n:n}$ are asymptotically independent. So are the order statistics $X_{1:n}$ and $Y_{1:n}$. Despite of this fact, the Baker's type BB distribution constructed on the base of the copula (24) and (25) has enough large correlation. Indeed, consider Baker's type BB distribution (9) with the "kernel" copula (24) and with the "kernel" copula (25):

$$
\begin{aligned}
\check{G}_{+}^{(n)}(x,y) &= \frac{1}{n}\sum_{k=1}^{n}C_{n,n:n}(F_{X}^{k:n}(x),F_{Y}^{k:n}(y))\\[2mm]
&= \frac{1}{n}\sum_{k=1}^{n}F_{X}^{k:n}(x)F_{Y}^{k:n}(y)\left[1+\alpha(1-\left[F_{X}^{k:n}(x)\right]^{1/n})(1-\left[F_{Y}^{k:n}(y)\right]^{1/n})\right]^{n}\\[2mm]
\hat{G}_{+}^{(n)}(x,y) &= \frac{1}{n}\sum_{k=1}^{n}C_{1,n:n}(F_{X}^{k:n}(x),F_{Y}^{k:n}(y))\\[2mm]
&= \frac{1}{n}\sum_{k=1}^{n}\Big\{F_{X}^{k:n}(x)+F_{Y}^{k:n}(y)-1+\Big[(1-F_{X}^{k:n}(x))^{1/n}+(1-F_{Y}^{k:n}(y))^{1/n}-1\\[2mm]
&\quad+(1-(1-F_{X}^{k:n}(x))^{1/n})(1-(1-F_{Y}^{k:n}(y))^{1/n})\\[2mm]
&\quad\times\{1+\alpha(1-F_{X}^{k:n}(x))^{1/n}(1-F_{Y}^{k:n}(y))^{1/n}\}\Big]^{n}\Big\}.
\end{aligned}
$$

For example, the correlation coefficient of $\check{G}_{+}^{(n)}(x,y)$ for $Uniform(0,1)$ marginals is $\rho_{\hat{G}_{+}^{(n)}}=0.4985$ for $n=2$.

3.2 The Baker's type BB distributions with "kernel" copula of bivariate Gumbel's extreme order statistics

Let

$$
C(u,v)=\frac{uv}{1+u-uv}. \tag{26}
$$

(26) is the copula of the Gumbel's bivariate logistic distribution function

$$
H_{X,Y}(x,y)=(1+e^{-x}+e^{-y})^{-1},\; x\geq 0, y\geq 0
$$

with standard logistic marginal distributions $H_X(x) = (1+e^{-x})^{-1}$ and $H_Y(y) = (1+e^{-y})^{-1}$, $x \geq 0, y \geq 0$. (Gumbel (1961), see also Nelsen (2005), page 28). Using

$$F(x,y) = \frac{F_X(x)F_Y(y)}{F_X(x) + F_Y(y) - F_X(x)F_Y(y)},$$

one obtains from (20)

$$C_{n:n}(t,s) = \frac{ts}{(t^{1/n} + s^{1/n} - t^{1/n}s^{1/n})^n} \tag{27}$$

and from (21)

$$
\begin{aligned}
C_{1:n}(t,s) \;=\; & t+s-1+ \\
& + \Big[(1-t)^{1/n} + (1-s)^{1/n} - 1 \\
& + \frac{(1-(1-t)^{1/n})(1-(1-s)^{1/n})}{2 - (1-t)^{1/n} - (1-s)^{1/n} - (1-(1-t)^{1/n})(1-(1-s)^{1/n})} \Big]^n .
\end{aligned}
\tag{28}
$$

It is seen that

$$\lim_{n\to\infty} C_{n:n}(t,s) = \lim_{n\to\infty} \frac{ts}{(t^{1/n} + s^{1/n} - t^{1/n}s^{1/n})^n} = ts = \Pi(t,s).$$

4 Joint distribution of bivariate order statistics for Frechét upper bound copula

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be a bivariate sample with joint distribution function $F(x,y) = C(F_X(x), F_Y(y))$. In this section our aim is firstly, to investigate the joint distribution function of bivariate order statistics $(X_{r:n}, Y_{r:n})$ for a copula with maximal correlation, i.e. the Frechéet upper bound. We are interested then in distribution function

$$K_+^{(n)}(x,y) = \frac{1}{n} \sum_{r=1}^{n} \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k=a}^{b} c(n,k;i,j) p_{11}^k p_{12}^{i-k} p_{21}^{j-k} p_{22}^{n-i-j+k},$$

in the case where the marginal distributions are uniform and $C(u,v) = \min(u,v)$. Recall that the coefficients $c(n,k;i,j)$ and $p_{11}, p_{12}, p_{21}, p_{22}$ are given in (6),(5). Secondly, we consider a distribution introduced in (9) with uniform marginals, i.e.

$$
\begin{aligned}
G_+^{(n)}(x,y) \;&=\; \frac{1}{n} \sum_{k=1}^{n} C(F_X^{k:n}(x), F_Y^{k:n}(y)) \\
&=\; \frac{1}{n} \sum_{r=1}^{n} C(F_{U_{r:n}}(x), F_{V_{r:n}}(y)),
\end{aligned}
$$

11

where $C(t, s) = \min(t, s)$ and $F_{U_{r:n}}(x) = \sum_{i=r}^{n} \binom{n}{i} x^i (1 - x)^{n-i}$ and $F_{V_{r:n}}(y) = \sum_{i=r}^{n} \binom{n}{i} y^i (1 - y)^{n-i}$.

Assume that marginal distributions are $Uniform(0, 1)$, i.e. $F_X(x) = x, 0 \leq x \leq 1$, $F_Y(y) = y$, $0 \leq y \leq 1$. Denote by $(U_i, V_i), i = 1, 2, ..., n$ the random sample from the bivariate distribution $C(u, v), 0 \leq u \leq 1, 0 \leq v \leq 1$ and $U_{1:n} \leq U_{2:n} \leq \cdots \leq U_{n:n}, V_{1:n} \leq V_{2:n} \leq \cdots \leq V_{n:n}$ be corresponding order statistics. Then from (4) for $r = s$ one has

$$
\begin{aligned}
P\{U_{r:n} &\leq u, V_{r:n} \leq v\} \\
&= \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k=a}^{b} c(n, k; i, j) C^k(u, v)(u - C(u, v))^{i-k} \\
&\quad \times (v - C(u, v))^{j-k} (\bar{C}(u, v))^{n-i-j+k}, \ 0 \leq u \leq 1, 0 \leq v \leq 1.
\end{aligned}
\tag{29}
$$

and

$$
K_+^{(n)}(u, v) = \frac{1}{n} \sum_{r=1}^{n} \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k=a}^{b} c(n, k; i, j) C(u, v)^k (u - C(u, v))^{i-k}
$$

$$
(v - C(u, v))^{j-k} (\bar{C}(u, v))^{n-i-j+k}, 0 \leq u \leq 1, 0 \leq v \leq 1.
\tag{30}
$$

Lemma 1. The joint distribution of $U_{r:n}$ and $V_{r:n}$ can be represented as

$$
\begin{aligned}
P\{U_{r:n} &\leq u, V_{r:n} \leq v\} \\
&= \sum_{i=r}^{n} \binom{n}{i} C^i(u, v) \left[v - C(u, v) + \bar{C}(u, v)\right]^{n-i} \\
&\quad + \sum_{j=r}^{n} \binom{n}{j} C^j(u, v) \left[u - C(u, v) + \bar{C}(u, v)\right]^{n-j} \\
&\quad - \sum_{i=r}^{n} \binom{n}{i} C^i(u, v)(\bar{C}(u, v))^{n-i} \\
&\quad + \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k \neq i \neq j} c(n, k; i, j) C(u, v)^k (u - C(u, v))^{i-k} \\
&\quad \times (v - C(u, v))^{j-k} (\bar{C}(u, v))^{n-i-j+k}.
\end{aligned}
\tag{31}
$$

Proof. Separating terms for summation in (29) for $k = i = j, k = i \neq j$ and

12

$k = j \neq i$ we have

$$
\begin{aligned}
P\{U_{r:n} \;\; \leq \;\; & u, V_{r:n} \leq v\} \\
= \;\; & \sum_{i=r}^{n} c(n,i,i,i) C^i(u,v)(\bar{C}(u,v))^{n-i} \\
& + \sum_{i=r}^{n} \sum_{j=i+1}^{n} c(n,i,i,j) C^i(u,v)(u - C(u,v))^{i-i}(v - C(u,v))^{j-i}(\bar{C}(u,v))^{n-i-j+i} \\
& + \sum_{j=r}^{n} \sum_{i=j+1}^{n} c(n,i,i,j) C^j(u,v)(u - C(u,v))^{i-j}(v - C(u,v))^{j-j}(\bar{C}(u,v))^{n-i-j+j} \\
& + \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k \neq i \neq j} c(n,k;i,j) C(u,v)^k (u - C(u,v))^{i-k}(v - C(u,v))^{j-k}(\bar{C}(u,v))^{n-i-j+k}
\end{aligned}
$$

$$
\begin{aligned}
= \;\; & \sum_{i=r}^{n} \binom{n}{i} C^i(u,v)(\bar{C}(u,v))^{n-i} \\
& + \sum_{i=r}^{n} \sum_{j=i+1}^{n} \frac{n!}{i!(j-i)!(n-j)!} C^i(u,v)(v - C(u,v))^{j-i}(\bar{C}(u,v))^{n-j} \\
& + \sum_{j=r}^{n} \sum_{i=j+1}^{n} \frac{n!}{j!(i-j)!(n-i)!} C^j(u,v)(u - C(u,v))^{i-j}(\bar{C}(u,v))^{n-i} \\
& + \sum_{i=r}^{n} \sum_{j=r}^{n} \sum_{k \neq i \neq j} c(n,k;i,j) C(u,v)^k (u - C(u,v))^{i-k} \\
& \times (v - C(u,v))^{j-k}(\bar{C}(u,v))^{n-i-j+k}.
\end{aligned}
\tag{32}
$$

Consider the second summation in (32) and changing index in the inner sum as
$j - i = k$ $(j = i + k,\ j = i + 1 \Longrightarrow k = 1, j = n \Longrightarrow k = n - i)$ we have

$$
\begin{aligned}
& \sum_{i=r}^{n} \sum_{j=i+1}^{n} \frac{n!}{i!(j-i)!(n-j)!} C^i(u,v)(v - C(u,v))^{j-i}(\bar{C}(u,v))^{n-j} \\
= \;\; & \sum_{i=r}^{n} C^i(u,v) \frac{n!}{i!(n-i)!} \sum_{k=1}^{n-i} \frac{(n-i)!}{k!(n-i-k)!} (v - C(u,v))^k (\bar{C}(u,v))^{n-i-k} \\
= \;\; & \sum_{i=r}^{n} \binom{n}{i} C^i(u,v) \left[\sum_{k=0}^{n-i} \frac{(n-i)!}{k!(n-i-k)!} (v - C(u,v))^k (\bar{C}(u,v))^{n-i-k} - (\bar{C}(u,v))^{n-i} \right] \\
= \;\; & \sum_{i=r}^{n} \binom{n}{i} C^i(u,v) \left[v - C(u,v) + \bar{C}(u,v) \right]^{n-i} - \sum_{i=r}^{n} \binom{n}{i} C^i(u,v)(\bar{C}(u,v))^{n-i}. \tag{33}
\end{aligned}
$$

13

Analogously, the third term in (32) can be written as

$$\sum_{j=r}^{n}\sum_{i=j+1}^{n}\frac{n!}{j!(i-j)!(n-i)!}C^j(u,v)(u-C(u,v))^{i-j}(\bar{C}(u,v))^{n-i}$$

$$= \sum_{j=r}^{n}\binom{n}{j}C^j(u,v)\left[u-C(u,v)+\bar{C}(u,v)\right]^{n-i} \tag{34}$$

$$- \sum_{i=r}^{n}\binom{n}{i}C^i(u,v)(\bar{C}(u,v))^{n-i}.$$

Taking into account (33) and (34) in (32) we have (31).

Theorem 2. If $C(u,v) = \min(u,v)$, then

$$P\{U_{r:n} \leq u, V_{r:n} \leq v\}$$

$$= \begin{cases} \sum_{i=r}^{n}\binom{n}{i}u^i\left[1-u\right]^{n-i} & if \quad u \leq v \\ \sum_{i=r}^{n}\binom{n}{i}v^i\left[1-v\right]^{n-i} & if \quad u > v \end{cases}$$

$$= \begin{cases} P\{U_{r:n} \leq u\} & if \quad u \leq v \\ P\{V_{r:n} \leq v\} & if \quad u > v \end{cases}$$

Proof. Consider (31). Let $C(u,v) = \min(u,v)$. Then it is clear that $(u-C(u,v))^{i-k}(v-C(u,v))^{j-k} = 0$ for those k satisfying $k \neq i$ and $k \neq j$, because $u-C(u,v) = u-u = 0$ if $u \leq v$ and $v-C(u,v) = 0$ if $u > v$. Therefore the last term of (31) vanishes and we have from (31)

$$P\{U_{r:n} \leq u, V_{r:n} \leq v\}$$

$$= \begin{cases} \sum_{i=r}^{n}\binom{n}{i}u^i\left[1-u\right]^{n-i} & if \quad u \leq v \\ \sum_{i=r}^{n}\binom{n}{i}v^i\left[1-v\right]^{n-i} & if \quad u > v \end{cases} \tag{35}$$

$$= \begin{cases} P\{U_{r:n} \leq u\} & if \quad u \leq v \\ P\{V_{r:n} \leq v\} & if \quad u > v \end{cases}.$$

Consider now, the Baker's Type I BB copula obtained from (35)

$$K_{+}^{(n)}(u,v) = \frac{1}{n}\sum_{r=1}^{n}P\{U_{r:n} \leq u, V_{r:n} \leq v\}$$

$$= \begin{cases} \frac{1}{n}\sum_{r=1}^{n}\sum_{i=r}^{n}\binom{n}{i}u^i\left[1-u\right]^{n-i} & if \quad u \leq v \\ \frac{1}{n}\sum_{r=1}^{n}\sum_{i=r}^{n}\binom{n}{i}v^i\left[1-v\right]^{n-i} & if \quad u > v \end{cases}$$

14

$$= \begin{cases} u & if \quad u \le v \\ v & if \quad u > v \end{cases} = W(u,v) = \min(u,v)$$

is nothing but the Frechét upper bound itself. In other words the Baker's BB distribution with uniform marginals and underlying joint distribution being Frechét upper bound generates the Frechét upper bound.

Now consider the Baker's Type II BB distribution with the "kernel" copula being Frechét upper bound, i.e.

$$G_+^{(n)}(x,y) = \frac{1}{n} \sum_{r=1}^{n} C(F_{U_{r:n}}(x), F_{V_{r:n}}(y)),$$

where $C(t,s) = \min(t,s)$ and $F_{U_{r:n}}(x) = \sum_{i=r}^{n} \binom{n}{i} x^i (1-x)^{n-i}$ and $F_{V_{r:n}}(y) = \sum_{i=r}^{n} \binom{n}{i} y^i (1-y)^{n-i}$. It follows that

$$
\begin{aligned}
G_+^{(n)}(x,y) &= \frac{1}{n} \sum_{r=1}^{n} \min(F_{U_{r:n}}(x), F_{V_{r:n}}(y)) \\
&= \frac{1}{n} \sum_{r=1}^{n} \min(F_{U_{r:n}}(x), F_{V_{r:n}}(y)) \\
&= \begin{cases} \frac{1}{n}\sum_{r=1}^{n} \sum_{i=r}^{n} \binom{n}{i} x^i(1-x)^{n-i} & if \quad x \le y \\ \frac{1}{n}\sum_{r=1}^{n} \sum_{i=r}^{n} \binom{n}{i} y^i(1-y)^{n-i} & if \quad x > y \end{cases} \\
&= \min(x,y).
\end{aligned}
$$

Therefore, if in the constructions $K_+^{(n)}(x,y)$ and $G_+^{(n)}(x,y)$ with $Uniform(0,1)$ marginals one uses the Frechét upper bound copula $W(t,s)$, then obtained copula is again the Frechéet upper bound copula. From the Example 1 one observes that if in the same construction one uses FGM copula then different distributions with high correlation can be obtained. For different constructions satisfying conditions of the Example 1, we have the following set of inequalities for correlation coefficient

$$\rho_M \le \rho_{G_-^{(n)}} \le \rho_{K_-^{(n)}} \le \rho_{H_-^{(n)}} \le \rho_{H_+^{(n)}} \le \rho_{K_+^{(n)}} \le \rho_{G_+^{(n)}} \le \rho_W,$$

where ρ_M is the correlation coefficient of the Frechéet lower bound $M(t,s) = \max(t+s-1,0)$.

References

[1] B. Arnold, N. Balakrishnan and H.N. Nagaraja, *A First Course in Order Statistics* (1992) John Wiley & Sons.

[2] C. Amblard and S. Girard, Symmetry and dependence properties within a semiparametric family of bivariate copulas, *Nonparametric.Stat.* 14(6) (2002) 715-727.

[3] I. Bairamov and S. Kotz, Dependence structure and symmetry of Huang-Kotz FGM distributions and their extensions, *Metrika* 56 (1) (2002) 55-72.

[4] I. Bairamov and S. Kotz, On a new family of positive quadrant dependent bivariate distributions, *International Mathematical Journal* 3(11) (2003) 1247–1254.

[5] I. Bairamov, S. Kotz and M. Bekçi, New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics, *Journal of Applied Statistics,* 28(5) (2001) 521–536.

[6] I. Bairamov, K. Bayramoglu. From Huang-Kotz distribution to Baker's distribution, *J. Multivariate Anal.,2011, in press.*

[7] R. Baker, An order-statistics-based method for constructing multivariate distributions with fixed marginals, *J. Multivariate Anal.* 99 (2008) 2312 2327.

[8] N. Balakrishnan and C.D. Lai, *Continuous Bivariate Distributions.* Second Edition, (2009) Springer

[9] H. David, *Order Statistics.* 2 nd edn. (1981) John Wiley, New York.

[10] M. Fischer, I. Klein, Constructing generalized FGM copulas by means of certain univariate distributions, *Metrika* 65 (2007) 243 260.

[11] M. Fréchet, On correlation matrices with fixed margins (Sur les tableaux de corrélation dont les marges sont données.), *Ann. Univ. Lyon Sect.* A 14 (1951) 53-77.

[12] W. Hoeffding, Scale-invariant correlation theory. (Maszstabinvariante Korrelationstheorie.), *Schr. Math. Inst. Univ. Berlin* 5 (1940) 181 233.

[13] J.S. Huang and S. Kotz, Modifications of the Farlie Gumbel Morgenstern distributions. A tough hill to climb, *Metrika* 49 (1999) 135 -145.

[14] G.D. Lin and J.S. Huang, A note on the maximum correlation for Baker's bivariate distributions with fixed marginals.*J. Multivariate Anal.* 101 (2010) 2227 2233

[15] R. Nelsen *An Introduction to Copulas.* Second Edition. (2005) Springer.